

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-105235

(43) Date of publication of application : 21.04.1995

(51)Int.Cl.

G06F 17/30

G10L 3/00

G10L 5/06

(21)Application number : 05-252798

(71)Applicant : SHARP CORP

(22)Date of filing : 08.10.1993

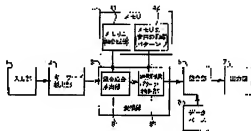
(72)Inventor : MATSUOKA YASUKO

## (54) METHOD AND DEVICE FOR IMAGE RETRIEVAL

(57)Abstract:

**PURPOSE:** To lighten the burden of image feature extraction, etc., on an operator at the time of data base input by extracting an imitation sound component which relates to a key word from a key word which is inputted for retrieval and generating the frequency pattern of the speech corresponding to the imitation sound component.

**CONSTITUTION:** The key word to be retrieved is extracted from an input sentence, and imitation sound information which features the key word is made to correspond to this key word. For each piece of imitation sound information, a 1st memory 41 is stored with an imitation sound component key word audience, a goal post, a ball, and a whistle as imitation sound components in the case of, for example, soccer. The imitation sound component extraction part 31 of a conversion part 3 extracts imitation sound components by referring to the 1st memory 41. The imitation sound components which are thus extracted are sent to the standard frequency pattern extraction part 3 and frequency patterns corresponding to the respective imitation sound component keywords of the audience, goal post, ball, and whistle are read out of a 2nd memory 42 by referring to the memory 42.



## LEGAL STATUS

[Date of request for examination]

04.07.1997

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

3021252

[Date of registration]

14.01.2000

[Number of appeal against examiner's decision]

of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

特開平7-105235

(43) 公開日 平成7年(1995)4月21日

(51) Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/30				
G 1 0 L 3/00	5 3 1 N	9379-5H		
5/06	A	9379-5H		
		9194-5L	G 0 6 F 15/ 40	3 7 0 G
		9194-5L	15/ 403	3 2 0 Z
審査請求 未請求 請求項の数2 O L (全 5 頁)				

(21) 出願番号 特願平5-252798

(22) 出願日 平成5年(1993)10月8日

(71) 出願人 000005049

シャープ株式会社  
大阪府大阪市阿倍野区長池町22番22号

(72) 発明者 松岡 靖子

大阪府大阪市阿倍野区長池町22番22号 シ  
ャープ株式会社内

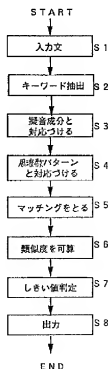
(74) 代理人 弁理士 梅田 勝

## (54) 【発明の名称】 画像検索方法及び画像検索装置

## (57) 【要約】

【目的】 画像と同期して記憶されている音声の中の擬音情報を検索キーとして用いることにより、データベース入力時の画像の特徴抽出等の作業者にかかる負担を軽減し、また、キーワード等の言語的な情報では表現しきれなかった画像の曖昧さを少なくする画像検索装置を提供することを目的とする。

【構成】 画像と音声とが同期して記録された画像を検索する方法において、検索用に入力されたキーワードからそのキーワードに関連した擬音成分を抽出すると共に、擬音成分に対応した画像データの特徴を表すような音声の周波数パターンを作成し、その周波数パターンを検索キーとして使用して画像検索を行う。



#### 【特許請求の範囲】

【請求項1】 画像と音声とが同期して記録された画像を検索する方法において、検索用に入力されたキーワードから該キーワードに関連した擬音成分を抽出すると共に、擬音成分に対応した音声の周波数パターンを作成し、該周波数パターンを検索キーとして使用することを特徴とする画像検索方法。

【請求項2】 あらかじめ画像と音声とが同期して記録されているデータベースと、入力文を入力する入力部と、入力された入力文から検索の対象となるキーワードを抽出するキーワード抽出部と、各キーワードに対応する擬音成分キーワード、及び各擬音成分キーワードに対応した音声の周波数パターンを持つメモリと、抽出したキーワードを擬音成分に対応づけると共に前記メモリから擬音成分に対応した音声の周波数パターンを抽出する変換部と、メモリから抽出した音声の周波数パターンとデータベースの音声データとのマッチングをとる整合部と、マッチングの結果から画像を出力する出力部とを具備してなることを特徴とする画像検索装置。

#### 【発明の詳細な説明】

##### 【0001】

【産業上の利用分野】本発明は、画像検索方法及び画像検索装置に係り、特に画像と音声とが同期して記録されたデータを検索するための画像検索方法及び画像検索装置に関するものである。

##### 【0002】

【従来の技術】放送局等においては、ビデオテープ等、画像と音声とを同期して記録した画像データが大量に蓄積されており、これらのデータから自分の必要とするデータを高速に取り出すことができる画像検索装置が望まれている。

【0003】従来の画像検索装置における画像検索方法には、あらかじめ画像に付与したキーワードにより検索を行うキーワード法と、画像自体の特徴抽出を行いその特徴によりマッチングを行って検索をする特徴抽出法がある。

【0004】前者は、各画像データに整理番号や関連のあるキーワードをつける方式であり、画像の持つ情報をキーワードで表し、キーワードどうしに関係を持たせることにより画像を検索しやすくしている。

【0005】後者は、画像自体を直接検索キーとして用いる方法であり、画像認識の手法を用いて画像自体の特徴抽出を行い、その特徴を検索キーとし、検索キーとマッチングした画像を出力する。

##### 【0006】

【発明が解決しようとする課題】しかしながら、前述のキーワード法は、画像に対するキーワードの付け方が人

により異なるので、検索者がキーワードの入力者と異なる場合に必要な画像を検索することが難しい。これは、画像情報には、キーワードのような言語的な情報で表し得る情報だけでなく、意味的な情報や感覚的な情報が含まれているので、キーワードだけでは表現できなかったり、適当なキーワードが見つからなかったりするためである。また、検索者は、検索したい画像をはっきりと指定する必要があるので、入力時に付けたキーワードを覚えておかなければならなかった。しかし、データベースが大きくなると、すべての画像についてキーワードを覚えておくことは困難になる。

【0007】また、前述の特徴抽出法の場合、画像からの特徴抽出とマッチングをどのように行うかが問題である。そして、画像の特徴は、人によってさまざまな見方があり、一瞬に決めることがず、検索者が、検索したい画像をどのように入力するのかも問題である。特に、検索したい画像のイメージがあいまいな場合には、装置が検索者の意図を読み取らなければならない。また、いずれの場合でも、画像をデータベースに入力するときに手間がかかった。

【0008】本発明は、このような点に鑑みなされたものであり、画像と同期して記憶されている音声の中の擬音情報を検索キーとして用いることにより、データベース入力時の画像の特徴抽出等の作業者にかかる負担を軽減し、また、キーワード等の言語的な情報では表現しきれなかった画像の曖昧さを少なくする画像検索装置を提供することを目的とする。

##### 【0009】

【課題を解決するための手段】上記目的を達成するために、本出願の請求項1に記載の発明では、画像データの特徴を表すような音声の周波数パターンを検索キーとして利用して画像検索を行なうことを特徴とする。

【0010】さらに、本出願の請求項2に記載の発明では、音声周波数パターンを検索キーとして使用できる検索装置として、あらかじめ画像と音声とが同期して記録されているデータベースと、入力文を入力する入力部と、入力された入力文から検索の対象となるキーワードを抽出するキーワード抽出部と、各キーワードに対応する擬音成分キーワード、及び擬音成分に対応する音声の周波数パターンを持つメモリと、抽出したキーワードを擬音成分に対応づけると共に前記メモリから擬音成分に対応した音声の周波数パターンを抽出する変換部と、メモリから抽出した音声の周波数パターンとデータベースの音声データとのマッチングをとる整合部と、マッチングの結果から画像を出力する出力部とを具備することを特徴とする。

##### 【0011】

【作用】本出願の請求項1に記載の発明によれば、検索用に入力されたキーワードから該キーワードに対応する擬音成分を抽出するとともに、この擬音成分に対応した

音声の周波数パターンを抽出しこれをも検索キーとして用いるため、よりきめの細かい画像検索を行なうことができる。

【0012】また、本出願の請求項2に記載の発明によれば、キーボード等の入力部から入力された入力文は、キーワード抽出部においてキーワードが抽出され、変換部において擬音成分に対応した音声の周波数パターンを持つメモリから、抽出した擬音成分に対応した音声の周波数パターンが抽出される。その後、整合部においてメモリ部から抽出された各擬音成分の音声の周波数パターンとデータベースの音声パターンとのマッチングがとられ、出力部からマッチングした音声パターンと同期する画像が出力されることになるので、キーボードから文を入力するだけで、データベースに記録された音声をも検索に利用することができ、効率的な検索が可能な検索装置を得ることができる。

【0013】

【実施例】以下、本発明の好適な実施例を図を参照しながら説明する。図1は、本発明にかかる画像検索装置の一実施例の全体構成図である。図1において、1は検索者が入力文を入力する入力部、2は入力された入力文から検索の対象となるキーワードを抽出するキーワード抽出部、3は抽出されたキーワードに対応する擬音成分を抽出すると共に、擬音成分に対応した音声波形の標準パターンを抽出する変換部であり、後述のメモリ4を参照してキーワードに対応する擬音成分を抽出する擬音成分抽出部31と、後述のメモリ4を参照して擬音成分に対応した音声周波数パターンを抽出する標準周波数パターン抽出部32より成る。そして、4はメモリであり、キーワードに対応した擬音成分キーワードを記憶する第1のメモリ41と、擬音成分によって特定される実際の音声の周波数パターンを記憶する第2のメモリ42によって構成されている。また、5は各擬音成分の周波数パターンとデータベースの音声データとのマッチングを取って両パターンの類似度をとり全ての類似度を合計し、しきい値を判定する整合部、6は検索対象となる画像と音声同期しているデータを蓄積しているデータベース、7は前記整合部のしきい値の判定結果を基に検索結果を出力する出力部である。

【0014】次に、図2のフローチャートを参照して、図1に係る画像検索装置の動作を説明する。図1のデータベース5には、あらかじめ画像と音声とが同期して記録されたデータが蓄えられている。検索者はまず、入力部1から入力文を入力する（S1）。入力の形式としては、キーボード、手書き入力、音声などさまざまなものがあるが、入力部1ではどれを用いても構わない。入力文から検索の対象となるキーワードが抽出される（S2）。

【0015】ここで、キーワードには、あらかじめキーワードを特徴づけるような擬音情報が対応づけられてい

る。この擬音情報とは、車の走る音や犬の鳴き声、人の歓声など、人の話す言葉とは異なる音のことである。画像と音声とが同期したデータとしては、例えばテレビ番組を録画したビデオがあるが、このデータの特徴を表すには、画像情報とともに音声情報も重要である。また、同じキーワードによって表わされる画像は、音声情報にも共通の情報があるので検索の際には、それらのうち擬音情報をも利用する。例えば、キーワードが「サッカー」の場合、サッカーに共通の擬音情報は、「観客の声」「ゴールの音」「ボールを蹴る時の音」「笛の音」等が考えられる。

【0016】このような各擬音情報に応じ、第1のメモリ41に、あらかじめ、例えば「サッカー」に対する擬音成分として、擬音成分キーワード「観客」、「ゴール」、「ボール」「笛」を記憶させておくものとする。すると、入力部1への入力文から抽出されたキーワードが「サッカー」であれば、変換部3の擬音成分抽出部31は、第1のメモリ41を参照して、擬音成分として前述の擬音成分キーワード「観客」、「ゴール」、「ボール」、「笛」を抽出する。

【0017】このように抽出された擬音成分は、変換部3の標準周波数パターン抽出部32に送られる。擬音成分に対応するデータは、音なので特有の音声波形を持つ。それを周波数分析することによって、周波数パターンが得られる。標準周波数パターン抽出部32は、第2のメモリ42を参照し、このメモリ42から各擬音成分キーワード「観客」、「ゴール」、「ボール」、「笛」に対応した周波数パターンを読み出してくる。

【0018】図3は、抽出されたキーワード8が「サッカー」の場合、「サッカー」に対して擬音成分キーワード9～12が対応づけられ、さらに各擬音成分キーワード9～12に応じて周波数パターン13～16が対応づけられている場合を図示している。図3の周波数パターン13～16において、パターンの横軸は時間を表し、縦軸は周波数を表している。

【0019】キーワードとして「サッカー」の場合を例にとって説明すれば、「サッカー」に対応する擬音成分キーワード9、10、11、12は、各々「観客」、「ゴール」、「ボール」、「笛」であり、抽出された周波数パターン13～16は、各々「サッカー場における観客の声」の代表的な周波数パターン、「サッカーボールを蹴る音」の代表的な周波数パターン、「サッカー場における笛の音」の代表的な周波数パターンである。

【0020】このようにして、変換部3では、キーワードから擬音成分キーワードへの対応づけを行ない（S3）、さらに、擬音成分キーワードからその擬音成分に対応する周波数パターンへの対応づけを行なう（S4）。

【0021】次に、検索者が入力した入力文について検索キーとなる周波数パターンが得られたら、整合部5において、各周波数パターンごとにデータベース6の音声情報中の音声データとマッチングをとる(S5)。ここで、画像と音声とが同期したデータのうち、音声データのみを検索に用い、音声データを周波数分析し、これをデータベースの入力パターンとする。

【0022】ここで、図4を用いて、データベースの入力パターン及び擬音成分パターンの類似度を求める方法について説明する。図4において、21はデータベースの入力パターンであり、検索者が入力した入力文における擬音成分の周波数パターン23と同様に、横軸が時間を表し、縦軸が周波数を表している。図4に示すように、データベースからの入力パターン21において、マッチング区間の先頭を語頭22とする。語頭22を定めて、擬音成分パターン23と入力パターン21との類似度を、例えばDPマッチングを用いて求めたとしたときの結果を、図4中のグラフで示し、横軸に時間、縦軸に類似度を取っている。ここでいう類似度とは、検索キーとなる擬音成分の周波数パターンと、図1におけるデータベース6の音声データの周波数パターンである入力パターンとがどれだけ類似しているかを表すものである。

【0023】例えばDPマッチングでは、同じ擬音パターンでも時間的に収縮している場合が考えられるので、語頭22を定めたら、マッチング区間24の長さを少し伸び縮みさせる。伸び縮みさせた区間ごとにくっつく類似度を求め、最もパターンの類似度の高いものをその語頭での類似度と決める。図4に示したような場合では、マッチング区間aとマッチング区間bとの両方で類似度を求める。語頭を時間軸方向に少しずつずらしていったら、データベースの入力パターン全体とマッチングすることができる。このような処理を、メモリ部4から得られた検索に必要なすべての擬音成分の周波数パターンについて行ない、周波数パターンごとの類似度を求め、グラフで示す。

【0024】なお、本実施例では、音声データの類似度を求める方法として、DPマッチングを用いたが、これに限定されるものではなく、この他にも多数あり、音声認識等に利用されている技術が応用できることはいうまでもない。

【0025】次に、上記のようにして求めた類似度のグラフを計算する(S6)。図4における類似度のグラフを初め、全ての類似度のグラフを加算したものが、図5に示すものである。次に、しきい値の判定を行う(S7)。これは、図5に示すしきい値25より類似度の高いグラフの部分が、必要な擬音成分を多く含むと考えられるからである。図5においてしきい値25より類似度が高い部分を、音声データに関する検索結果とする。データベースは、画像と音声とが同期しているから、音声と

合った画像を検索し、画像と音声と同時に出力する(S8)。但し、このしきい値は必要に応じて変更することができる。

【0026】以上の説明では、キーワード自体による検索は述べていないが、このキーワード自体をも利用した検索を合わせ持つように構成しても良いということはいうまでもない。

【0027】

【発明の効果】以上説明したように、本発明の画像検索方法及び画像検索装置によれば、データベース入力時にキーワードを付与する必要がなく、画像の特徴抽出も行わずによいので作業者の負担を軽減することができる。これまで画像検索を難しくしていた画像のあいまいさを擬音情報を使うことで補うことができ、また、同じキーワードであれば違ふ画像であっても検索することができる。これにより検索者は効率的な検索を行なうことができる。

【0028】また、動画像情報に比べ情報量の少ない音声情報を用いることによって、検索の手間を軽減することができる。さらに、各擬音成分にはそれぞれその擬音成分に対応する周波数パターンを持たせておきキーワードは異なっても同じ擬音成分を含んでいれば同じ周波数パターンを用いることができるので、メモリの削減になる。

【図面の簡単な説明】

【図1】本発明の一実施例の画像検索装置の全体構成図である。

【図2】図1の画像検索装置の画像処理の動作を説明するフローチャートである。

【図3】キーワードと擬音成分の周波数パターンとの対応づけを説明する図である。

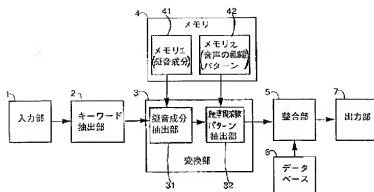
【図4】入力パターン及び擬音成分パターンの類似度を求める方法について説明する図である。

【図5】入力パターン及び擬音成分パターンの類似度による検索結果を説明する図である。

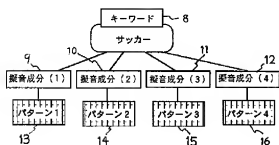
【符号の説明】

- 1 入力部
- 2 キーワード抽出部
- 3 変換部
- 4 メモリ
- 5 整合部
- 6 データベース
- 7 出力部
- 8 キーワード
- 9, 10, 11, 12 擬音成分キーワード
- 13, 14, 15, 16 周波数パターン
- 21 入力パターン
- 23 擬音成分パターン

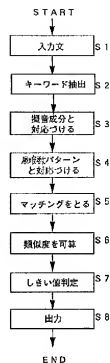
【図1】



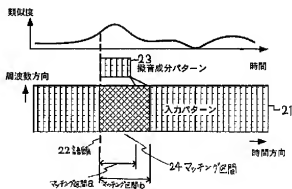
【図3】



【図2】



【図4】



【図5】

